ECONOMETRIC STRUCTURE OF PANEL DATA MODELS[1]

Phoebus J. Dhrymes[2]
Columbia University

January 2006

[2] This paper is an attempt to provide a concise and clear exposition of this topic.

1.1 Introduction

The study of empirical models based on cross-section times series data dates back well into the early 50s. However, the first modern attempt to consistently model the behavior of agents in such contexts can be traced to Balestra and Nerlove (1966), hereafter referred to as BN, who studied the demand for natural gas **by state** in the US, over the period 1957-1962.

The subsequent development of the subject, however, arose in the context of studies of human populations by "labor" economists, for want of a better term.

A sample of observations is said to be a **cross section**, if it refers to a number of economic agents **at the same period of time**, or reasonably close to the same period. For example the study of family budgets based on consumer expenditure surveys, or the study of the "determinants" of wages based on a survey of population, are examples of use of cross sections, or cross sectional samples. On the other hand a sample of observations on a single or a group of agents **over time** is said to be a time series sample. For example a study of aggregate consumption, or investment, or exports etc over a period of time are instances of use of time series samples.[1] The term Panel (or Panel data) refers to a situation where (the same) a number of agents are observed over multiple periods of time, so that for instance if there are n individuals observed at time t and $t = 1, 2, \ldots, T$, then **for each t** we have information over the same n individuals. This is more generally referred to as a Balanced Panel, and the term Panel is also applied to situations in which, by attrition or otherwise, not all n individuals are observed over all T periods, but a substantial number is. The term **repeated cross sections** or **pooled samples** typically refer to situations in which there may be substantial differences in the composition of the cross sections, although broadly conforming to the original specification(s).

The purpose of this note is exposit in simple form the basic structure of these models, indicating clearly their relation to the general linear model broadly and, when n is fixed and T is large, to the "seemingly unrelated regressions" model (SUR) in particular.

It does not aim, at this stage, to be an exhaustive treatment of the subject.

[1]Over the last fifteen years or so the term time-series has acquired a very specialized meaning in econometrics connected to the study of stochastic processes. However, our use of the term **time series** here typically does not refer to this meaning.

1.2 Notation

Because the study of panel data models operates in **three dimensions**, the number of agents, the number of periods and the number of variables (dependent and independent in the regression sense of the terms), care is required in devising a notation to represent the data. We adopt the following conventions. In Eq. (1) below we represent the typical econometric relationship estimated in a panel setting,

$$y_{ti} = x_{ti\cdot}\beta + \gamma_i + u_{ti}, \quad i = 1, 2, \ldots, n, \quad t = 1, 2, \ldots T, \tag{1.1}$$

where n is the number of agents observed, so that it is the dimension of the cross section, T is the number of the periods of observation, so that it is the dimension of the time series, and $x_{ti\cdot}$ is **a k element row vector**, containing the independent or explanatory variables of the model. The term γ_i is an idiosyncratic, time invariant entity that refers exclusively to agent i, and u_{ti} is the standard error term. The term γ_i may be random (so that it is distributed over the agents through a time invariance distribution), or nonrandom, like a distinct constant term. [2]

Because of the three-dimensional nature of panel data, it is not possible to give a single matrix representation for the entire sample; we shall now define the entities in terms of which we shall represent the data.

$$Y = (y_{ti}), \ t = 1, 2, \ldots, T, \quad i = 1, 2, \ldots n, \tag{1.2}$$

is the matrix of the dependent variable. The i^{th} **column** of it is

$$y_{\cdot i} = (y_{1i}, y_{2i}, \ldots, y_{Ti})' \tag{1.3}$$

and contains **all** T observations on **the individual agent** i. Its t^{th} row is

$$y_{t\cdot} = (y_{t1}, y_{t2}, \ldots, y_{tn}), \tag{1.4}$$

and contains **all observations** on the n agents at time t, i.e. it represents the t^{th} cross section.

With the help of this notation, we may represent all observations in the sample in two ways. First, we can exhibit, *seriatim* all observations on the first agent, then the second agent and so on. The second is to present *seriatim* all

[2] In fact in many studies allowing for particular individual effects removes as much as 30-40% of the observed variation in the dependent variable.

the observation on the first cross section, then those of the second cross section and so on. To implement the first, let

$$X_i = (x_{ti\cdot}), \quad t = 1, 2, \ldots, T \tag{1.5}$$

note that it is a $T \times k$ matrix and that

$$y_{\cdot i} = X_i \beta + \gamma_i e_T + u_{\cdot i}, \quad i = 1, 2, \ldots, n \tag{1.6}$$

represents all (time series) observations for the i^{th} agent, where e_T is a column vector of size T, all of whose elements are one. To implement the second, let

$$y'_{t\cdot} = X_{(t)} \beta + \gamma + u'_{t\cdot}, \quad \gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n)', \tag{1.7}$$

where

$$X_{(t)} = \begin{bmatrix} x_{t1\cdot} \\ x_{t2\cdot} \\ \vdots \\ x_{tn\cdot} \end{bmatrix}, \tag{1.8}$$

so that $X_{(t)}$ is a matrix of order $n \times k$, and Eq. (1.7) exhibits the observations on all agents in the t^{th} cross section.

To round out the required notation we introduce the following two definitions.

Definition 1. Let A be a matrix of order $q \times s$. The operator *vec* is defined by

$$a = \text{vec}(A) = \begin{bmatrix} a_{\cdot 1} \\ a_{2\cdot} \\ \vdots \\ a_{\cdot s} \end{bmatrix}, \tag{1.9}$$

i.e. it is of dimension $q \cdot s \times 1$ (a column vector) that exhibits the **columns** of A, *seriatim.*

Definition 2. Let A be a matrix as in Definition 1; the operator *rvec* is defined by

$$\text{rvec}(A) = \begin{bmatrix} a'_{1\cdot} \\ a'_{2\cdot} \\ \vdots \\ a'_{q\cdot} \end{bmatrix}, \tag{1.10}$$

i.e. it is of dimension $q \cdot s \times 1$ (a column vector) that exhibits **the transposed rows (so they become column vectors)** of A, *seriatim.*

Remark 1. Note that the two (column) vectors in Eqs. (1.9) and (1.10(**c ontainthe same elements** differently arranged.

If we put

$$Y = (y_{ti}), \quad U = (u_{ti}), \quad \text{both being } T \times n \text{ matrices}, \tag{1.11}$$

we can represent **the entire sample** in two ways; the first is

$$y^* = \mathrm{vec}(Y) = X^*\beta + \gamma \otimes e_T + u^*, \quad u^* = \mathrm{vec}(U), \quad X^* = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}, \tag{1.12}$$

while the second is

$$y = \mathrm{rvec}(Y) = X\beta + e_T \otimes \gamma + u, \quad u = \mathrm{rvec}(U), \quad X = \begin{bmatrix} X_{(1)} \\ X_{(2)} \\ \vdots \\ X_{(T)} \end{bmatrix}, \tag{1.13}$$

and both of them "look" like a general linear model (GLM).

1.3 Interpretation and Assumptions

1.3.1 Interpretation

The interpretation of the models in Eqs. (1.12), (1.13), depends on how we view the entities γ_i ; in the Balestra and Nerlove (BN) context, they relate to the **structural error** specification; thus, if we wish to estimate parameters by means of maximum likelihood (ML) methods, we need to state the distribution of the γ_i and u_{ti} , and obtain various moments such, e.g., the covariance matrix, so as to be able to write down the likelihood function, if normal, or the generalized least squares objective function. In particular, the conditional expectation of y_i given x is given by

$$E(y_i|x) = x_{ti\cdot}\beta, \tag{1.14}$$

because BN **assume** $E(\gamma_i|x) = 0$, $E(u_i|x) = 0$, for all i , and it is clear that the γ_i **are not** part of the mean specification.

In the context of such models as used in "labor" economics, the entities γ_i are viewed as part of the "mean component" of the dependent variable, and in fact they are a crucial component of their structure. The γ_i are presumed to denote native ability, which is not observed or observable; if it were, then it would have entered as one of the components of $x_{ti\cdot}$ and its coefficient would have been one of the β_i . For that reason it is entered as a single entity **denoting both** the "magnitude" of native ability as well as the manner of its impact on the dependent variable, i.e. its coefficient. From an operational point of view, in the context of the so called "within groups" estimation, it acts very much like a conventional regression **constant term**. A more sophisticated view holds the γ_i to be (proportional to) a random assignment of an **unobservable** property, such as ability, to agent i . In this context the entities γ_i , within the sample, are (proportional to) realizations of this random process.

An important consequence of this difference in interpretation between the BN and "labor" economics versions is that, in the first, the vector $x_{ti\cdot}$ is allowed to have a component which is **unity**, corresponding to the constant term of the regression, while in the second we assert

Convention 1. The vector β **does not contain** a constant term, i.e. the vector $x_{ti\cdot}$ does not contain a unit entry.

1.3.2 Assumptions

In the GLM we generally make three types of assumptions: (a) regarding the error process (b) regarding the explanatory or "independent" variables and (c) the relation between them.

In this literature, generally, not much is assumed about the explanatory variables, the vectors $x_{ti\cdot}$. The tendency is to consider the variables (y, x, γ, u) as being defined on a probability space, say (Ω , $\mathcal{A}$, $\mathcal{P}$), and subject to some joint distribution from which we can derive marginal and conditional distributions. They are also taken to be square integrable. In part the justification is that papers in this literature are seldom concerned with limiting distributions and take the distribution of the vector (y, x, γ, u) to be jointly normal. We shall depart from this practice, thus assuming in particular:

i. **Assumption 1** The explanatory variables, $x_{ti\cdot}$ are a realization of a square integrable stochastic process, so that $Ex_{ti\cdot}$ and $\mathrm{Cov}(x_{ti\cdot})$ are both finite and $A_{ij} = \mathrm{plim}_{T\to\infty} \frac{X_i' X_j}{T}$, exist for all i, j and for $i = j$ A_{ii} is nonsingular.

ii. **Assumption 2** The vectors $u_{\cdot i}$ are taken to be stationary, i.e. the marginal distribution[3] of $u_{t\cdot}$ has the property that for arbitrary q, s

$$f(u_{t\cdot}, u_{t+1\cdot}, \ldots, u_{t+q\cdot}) = f(u_{t+s\cdot}, u_{t+s+1\cdot}, \ldots, u_{t+s+q\cdot}),$$

are square integrable, and initially i.i.d. (independent identically distributed).

iii. **Assumption 3** The conditional expectations $E(u_{t\cdot}|x=0)$, (occasionally also $E(\gamma|x)=0$) and, moreover, their conditional (on x) covariance matrices are not a function of x and are as in their respective marginal distributions.

Remark 2. The conditions in Assumption 3, rely too heavily on the conventions of the GLM, especially as they relate to finite samples and requirements for the validity of the Gauss-Markov theorem. Unless considerably more structure is imposed, limiting distribution arguments would require that x and u are (statistically) independent.

Remark 3. The condition $E(u|x)=0$ is a bit stronger than $\mathrm{Cov}(x,u)=0$, whether we use the same time frame for both, or we are also claiming this property for $u_{t\cdot}$ and $x_{t'i\cdot}$, $t \neq t'$. To see this in the simplest possible context, let u, x be **scalar** random variables, each with marginal mean zero. Then,

$$\mathrm{Cov}(u,x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xuf(x,u)dxdu = \int_{-\infty}^{\infty} xf(x)\left(\int_{-\infty}^{\infty} uf(u|x)du\right)dx = 0 \tag{1.15}$$

because

$$h(x) = \int_{-\infty}^{\infty} uf(u|x)du = E(u|x) = 0, \quad \text{by assumption.}$$

The converse, however, is not true, i.e. $\mathrm{Cov}(u,x)=0$ does not necessarily imply that $E(u|x)=0$. This is so because

$$\mathrm{Cov}(u,x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xuf(x,u)dxdu = \int_{-\infty}^{\infty} xf(x)\left(\int_{-\infty}^{\infty} uf(u|x)du\right)dx$$

$$= \int_{-\infty}^{\infty} xh(x)f(x)dx = 0 \tag{1.16}$$

does not necessarily imply that $h(x)$ is equal to zero!

[3]The notation $f(\cdot)$ is used generically to denote a density, which may be marginal, joint, or conditional as the context requires.

1.4 Estimation

1.4.1 Within Groups Estimation

The estimation of panel data models involves basically the same procedures one employs in the estimation of general linear models, or systems of general linear models, with the added complication of the unobservable "ability", termed in this literature **unobserved heterogeneity**.

The term "within groups" is something of a misnomer and derives from an older usage in the context of analysis of variance, where one could define variance within groups and between groups; such considerations, however, are irrelevant in this case and the term survives only as a historical relic.

If we look at Eqs. (1.6), (1.12) we see that even though our primary interest is the parameter β, estimation by simple methods is hampered by the presence of the unobservable γ_i, for the i^{th} agent. If $T \geq 2$ we can, in fact, estimate β from **centered** data. We shall first consider the case where n is fixed and $T \to \infty$, recalling that initially we deal with the i.i.d. case.

Let $I_T - ee'/T$ be the (sample mean) centering matrix, Dhrymes (1978), pp. 18ff., where e is a column vector of T unities, we may transform Eq. (1.6) to obtain

$$\left(I_T - \frac{ee'}{T}\right) y_{\cdot i} = \left(I_T - \frac{ee'}{T}\right) X_i \beta + \left(I_T - \frac{ee'}{T}\right) u_{\cdot i}, \tag{1.17}$$

because the centering matrix and e_T are mutually orthogonal, i.e.

$$\left(I_T - \frac{ee'}{T}\right) e_T = 0.$$

While this operation has gotten rid of one problem, it has created a new one in that the covariance matrix of the error vector in Eq. (1.17) is no longer scalar. In fact, we also have another "problem" in that the sum of all equations therein is identically zero, indicating that the covariance matrix is **singular**. Note that the centering matrix is a **symmetric idempotent matrix** of rank $T-1$. Hence, see Dhrymes (2000), pp. 77-78, it has the decomposition

$$I_T - \frac{ee'}{T} = Q \begin{bmatrix} I_{T-1} & 0 \\ 0 & 0 \end{bmatrix} Q' = Q_1 Q_1', \tag{1.18}$$

where Q is the orthogonal matrix of the characteristic vectors, and Q_1 is (its) the sub-matrix corresponding to the **nonzero (unit)** roots.

Now, using another old device from Dhrymes (1969), transform Eq. (1.17) to obtain

$$Q_1' \left(I_T - \frac{ee'}{T}\right) y_{\cdot i} = Q_1' \left(I_T - \frac{ee'}{T}\right) X_i \beta + Q_1' \left(I_T - \frac{ee'}{T}\right) u_{\cdot i}. \tag{1.19}$$

Noting that $Q_1'(I_T - \frac{ee'}{T}) = Q_1'$, it is apparent that this entire operation could have been done *ab initio*; it was done in an extensive fashion only to clarify what is behind this transformation.

The virtue of this transformation is that it eliminates the parameter γ_i, by centering observations about sample means, and retains the property of the covariance matrix of the error as a **scalar** matrix of the form $\sigma_{ii} I_{T-1}$.

Since in the transformed context of Eq. (1.19), the entities therein obey the conditions for the Gauss-Markov theorem, see Dhrymes (1978), it follows that given X_i the estimator

$$\hat{\beta} = [X_i' Q_1 Q_1' X_i]^{-1} [X_i' Q_1 Q_1' y_{\cdot i}] = \beta + [(X_i' Q_1 Q_1' X_i]^{-1} [X_i' Q_1 Q_1' u_{\cdot i}], \quad (1.20)$$

is the best linear unbiased estimator in the context of Eq. (1.19).

If we consider the analogously transformed Eq. (1.12), we are dealing with

$$(I_n \otimes Q_1') y^* = (I_n \otimes Q_1') X^* \beta + (I_n \otimes Q_1') u^*, \quad (1.21)$$

because $(I_n \otimes Q_1')(\gamma \otimes e_T) = 0$. We also observe that

$$\mathrm{Cov}(u^*) = \Sigma \otimes I_T, \text{ so that } \mathrm{Cov}[(I_n \otimes Q_1') u^*] = \Sigma \otimes I_{T-1}. \quad (1.22)$$

Applying the same device as in Dhrymes (1969)[4] we pre-multiply by $\Sigma^{-(1/2)} \otimes I_{T-1}$ to obtain the final estimating form as

$$(\Sigma^{-1/2} \otimes Q_1') y^* = (\Sigma^{-1/2} \otimes Q_1') X^* \beta + (\Sigma^{-1/2} \otimes Q_1') u^*, \quad (1.23)$$

the covariance matrix of whose error term is $I_{n(T-1)}$. Thus, when Σ is known the least squares estimator in the context of the transformed model as in Eq. (1.23) is the "GMM", or optimal, estimator. If Σ is **not known** but can be estimated consistently, the estimator above is the GMM or optimal estimator in the sense that it has the smallest variance within the class of consistent estimators with "instruments" X^*.

As of now, we have not specified what Σ is. In the large T, fixed n context we have several (three) possibilities for

$$\Sigma = (\sigma_{ij}), \quad (1.24)$$

[4]The motivation for this device is to transform the equation whose parameters we are interested in estimating so that, at least asymptotically, it obeys the conditions of a Gauss-Markov like theorem. This idea (*mutatis mutandis*) is behind all subsequent developments in optimal non-linear simultaneous equations and generalized method of moments (GMM) procedures.

(a): no restrictions; (b): $\sigma_{ij} = 0, i \neq j$; and (c): the condition in (b) and in addition $\sigma_{ii} = \sigma^2, \ i = 1, 2, \ldots, n$.

In case (c), we have the typical condition imposed when $T = 1$ and a least squares procedure is applied. With $T \geq 2$, however, this condition is **unnecessarily** restrictive and we can operate instead with condition (b), which preserves the traditional assumption of independence in the error terms corresponding to different agents. At the same time, from Eq. (1.26) below, we can estimate consistently the unknown elements of Σ , thus rendering the estimator in Eq. (1.27) **possible**. The same is true if we operate with conditions a:, except that here we are asserting that the error terms for different agents are not independent, which is not an assumption made in this literature. Thus, in all cases with fixed n and large T it is possible to estimate consistently the σ_{ij} and hence the estimator of Eq. (1.23) is **feasible**. This shows that in the unestricted cse panel data efficien estimation is simply SUR estimation.

To complete the estimation phase we need to obtain an estimator for γ_i . Noting that in Eq. (1.6) $\gamma_{\cdot i}$ is dealt with as if it were the constant term of a standard GLM (general linear model), we can estimate

$$\hat{\gamma}_i = \bar{y}_i - \bar{x}_i \hat{\beta}, \quad i = 1, 2, \ldots, n, \tag{1.25}$$

where the symbols with overbars denote the sample means of the corresponding vectors ($y_{\cdot i}$), or matrices (X_i).

To enable us to estimate the efficient estimator for the system as a whole we can estimate the covariance matrix of $u_{t\cdot}$ as

$$\hat{\sigma}_{ij} = \frac{1}{T}(y_{\cdot i} - \hat{\gamma}_i e_T - X_i \hat{\beta})'(y_{\cdot j} - \hat{\gamma}_j e_T - X_j \hat{\beta}), \tag{1.26}$$

so that the efficient estimator of β is given by

$$\tilde{\beta} = [X^{*\prime}(\hat{\Sigma}^{-1} \otimes Q_1 Q_1')X^*]^{-1} X^{*\prime}(\hat{\Sigma}^{-1} \otimes Q_1 Q_1')y^* \tag{1.27}$$

Remark 4. Evidently the estimator in Eq. (1.27) is the only relevant one. The one in Eq. (1.20) is a “first stage” which serves to explain the procedure and estimates γ_i and $\sigma_{ij}, \ i, j = 1, 2, \ldots, n$. Notice also that the system estimator in Eq. (1.27), in the case of large T fixed n panels, should enable us to test for a more general form of heterogeneity. Thus, if we endow each agent, or groups of agents based on certain characteristics, with a **different parameter vector**, say, $\beta_{\cdot j}$, we may test the hypothesis that all such vectors are the same.

Otherwise we take, arbitrality, parameter homogeneity across agents or groups of agents as a **given** and, **even though the data permit it** we do not test this hypothesis (assumption).

Consistency of the estimators in Eqs. (1.25),(1.26), (1.27) is immediate by virtue of the assumption that the x's form a square integrable process, Assumption A1 and Assumption A3.

Not sufficiently precise assumptions have been made to ensure that the limiting distribution of these estimators exist, but with little additional loss of generality we can ensure that the Lindeberg condition holds, see Dhrymes (1989), pp.. This will permit us to conclude that

$$\sqrt{T}(\tilde{\beta}-\beta) \xrightarrow{\mathrm{d}} N(0,\Omega), \quad \Omega = \operatorname*{plim}_{T\to\infty}\left(\frac{X^{*\prime}(\hat{\Sigma}^{-1}\otimes Q_1Q_1^{\prime})X^{*}}{T}\right)^{-1}. \tag{1.28}$$

With a little additional effort we can devise a test for the hypothesis

$H_0 : \gamma_j = \gamma_1$
as against the alternative $H_1 : \gamma_i \neq \gamma_j$, for $i \neq j$.

We now take up the case, in the context of T large n fixed, where the vector $u_{t\cdot}$ is **not i.i.d.** but is either strictly stationary or follows some mixing distribution. This means that

$$\mathrm{Cov}(u_{t\cdot}^{\prime}) = K(t,t), \quad \mathrm{Cov}(u_{t\cdot}^{\prime},\ u_{t^{\prime}\cdot}^{\prime}) = K(t,t^{\prime}). \tag{1.29}$$

For example, if the process is strictly stationary $K(t,t) = K_0$; if it is **covariance stationary**, $K(t,t^{\prime}) = K(\tau)$, $\tau = t - t^{\prime}$, and $K(\tau) = K(-\tau)^{\prime}$. Thus, in the popular first order **stable** autoregression, AR(1), which is strictly stationary (as well as covariance stationary)

$$u_{t\cdot}^{\prime} = Ru_{t-1\cdot}^{\prime} + \epsilon_{t\cdot}^{\prime}, \quad u_{t\cdot}^{\prime} = \sum_{j=0}^{\infty} R^{j}\epsilon_{t-j\cdot}^{\prime}, \tag{1.30}$$

where the ϵ process is i.i.d. with Cov($\epsilon_{t\cdot}^{\prime}$) $= \Sigma$, we find

$$\mathrm{Cov}(u_{t\cdot}^{\prime}) = K_0 = \sum_{j=0}^{\infty} R^{j}\Sigma R^{\prime j}, \ \mathrm{Cov}(u_{t\cdot}^{\prime},\ u_{t-\tau}^{\prime}) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} R^{i}E(\epsilon_{t\cdot}^{\prime}, \epsilon_{t-\tau\cdot})R^{\prime j} = R^{\tau}K_0, \tag{1.31}$$

for **positive** τ , and $K_0R^{\prime|\tau|}$, for **negative** τ . Thus, the covariance matrix of the vector u in Eq. (1.13), for the stable first order autoregression above, is given by

$$\mathrm{Cov}(u) = \begin{bmatrix} K_0 & RK_0 & R^2K_0 & \dots & R^{T-1}K_0 \\ K_0R' & K_0 & RK_0 & \dots & R^{T-2}K_0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ K_0R'^{T-1} & K_0R'^{T-2} & K_0R'^{T-3} & \dots & K_0 \end{bmatrix} = K. \quad (1.32)$$

We shall now address the question of how to estimate the parameters of Eq. (1.21) when the error terms constitute an AR(1) process with covariance matrix as in Eq. (1.32)

In Eqs. (1.12) and (1.13) we have given two representations for the display of the observations on the model in Eq. (1.1). This was not done frivolously. Rather, the representation in Eq. (1.12) is most convenient for eliminating the "unobserved heterogeneity" parameters, γ_i , while the representation in Eq. (1.13) is most convenient when the error term follows a more complicated distribution such as for example the AR(1) considered just above; conversely Eq. (1.13) is most convenient in obtaining the covariance matrix of the error vector, when the latter is not i.i.d. but has a more complicated probabilistic structure, such as e.g. the AR(1) process noted above, but it is extremely inconvenient for the "within groups" estimator because it is not simple to eliminate the vector γ from Eq. (1.13) in any meaningful way.

Thus, we shall proceed by using Eq. (1.12) to eliminate the unobserved heterogeneity parameters, and we shall use Eq. (1.13) to obtain the covariance matrix of the error as we did in Eq. (1.32); having done that we must find some way to rearrange the elements of that matrix so that it becomes the covariance matrix of the vector as exhibited in Eq. (1.12). Noting that

$$u^* = \mathrm{vec}(U), \quad u = \mathrm{rvec}(U) \quad (1.33)$$

contain the same elements differently arranged, we find from Dhrymes (2000), pp. 137-139, that there exists a permutation matrix P_{nT} , such that

$$u^* = P_{nT}u, \quad \text{or} \quad u = P_{Tn}u^*, \quad P_{Tn} = \begin{bmatrix} I_n \otimes e'_{\cdot 1} \\ I_n \otimes e'_{\cdot 2} \\ \vdots \\ I_n \otimes e'_{\cdot T} \end{bmatrix}, \quad (1.34)$$

where $e_{\cdot t}$ is a T element **column vector** all of whose elements are zero except the t^{th} , which is unity.

Returning to Eq. (1.21) we thus find

$$\mathrm{Cov}[(I_n \otimes Q_1')u^*] = \Phi, \quad \text{where} \quad (1.35)$$

$$\Phi = S'KS, \quad S = \begin{bmatrix} I_n \otimes q_{1\cdot} \\ I_n \otimes q_{2\cdot} \\ \vdots \\ I_n \otimes q_{T-1\cdot} \end{bmatrix}, \tag{1.36}$$

and $q_{t\cdot}$ is the t^{th} row of Q_1. Notice that the rows of Q_1 have $T-1$ elements and thus $I_n \otimes q_{t\cdot}$ is $n \times n(T-1)$, so that S is $nT \times n(T-1)$; thus the matrix Φ is $n(T-1) \times n(T-1)$ and nonsingular, as required for the definition of the "GMM" estimator. The latter may be obtained as the OLS estimator in the context of

$$\Phi^{-(1/2)}(I_n \otimes Q_1')y^* = \Phi^{-(1/2)}(I_n \otimes Q_1')X^*\beta + \Phi^{-(1/2)}(I_n \otimes Q_1')u^*. \tag{1.37}$$

Given a consistent estimator of K (and thus $\hat{\Phi} = S'\tilde{K}S$) the feasible estimator is

$$\tilde{\beta} = [X^{*\prime}(I_n \otimes Q_1)\hat{\Phi}^{-1}(I_n \otimes Q_1')X^*]^{-1}X^{*\prime}((I_n \otimes Q_1)\hat{\Phi}^{-1}(I_n \otimes Q_1')y^*. \tag{1.38}$$

By arguments analogous to those leading to Eq. (1.28) we conclude that

$$\sqrt{T}(\tilde{\beta} - \beta) \xrightarrow{\mathrm{d}} N(0, \Phi^{-1}), \quad \Phi = \operatorname*{plim}_{T\to\infty} S'\tilde{K}S, \tag{1.39}$$

where $\tilde{K}$ is the consistent estimator of K obtained by **regressing** $\hat{u}_{t\cdot}$ on $\hat{u}_{t-1\cdot}$, thus obtaining $\hat{R}$, and by estimating

$$\hat{K}_0 = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_{t\cdot}'\hat{u}_{t\cdot}. \tag{1.40}$$

This procedure yields a **consistent estimator** because **we have assumed that the error process is a stable AR(1)**, and thus **strictly stationary**.

If the error process is more complex and/or is not stated parametrically, we can employ, *mutatis mutandis* a similar procedure by estimating the entries in the (general form of the) matrix $K = [K(t,t')]$, using the sample covariances, ($t \neq t'$)

$$\hat{K}(t,t') = \sum w(t,t')\hat{u}_{t\cdot}'\hat{u}_{t'\cdot}, \quad \hat{K}_0 = \frac{1}{T}\sum_{t=1}^{T} u_{t\cdot}'u_{t\cdot}, \tag{1.41}$$

provided stationarity is preserved, and where w is an appropriate weighting function, akin to the spectral windows (also known as kernels) employed when such problems are considered in the frequency domain.

When n is large and $T \geq 2$ is **fixed** our flexibility is considerably more circumscribed. This is so because, (a) no time series parameters can be consistently estimated, owing to the fixity of T, and (b) the cases

$$\Sigma = \mathrm{diag}(\sigma_{11}, \sigma_{22}, \ldots, \sigma_{nn}), \quad \Sigma = (\sigma_{ij}), \quad \sigma_{ij} \neq 0, \text{ for } i \neq j, \tag{1.42}$$

do not permit "efficient estimators", because the parameters in Σ cannot be consistently estimated. Thus we are reduced to assuming, implicitly, that $\Sigma = \sigma^2 I_n$ and obtaining OLS estimators. In the case of the **diagonal** covariance matrix in Eq. (1.42) we can consistently estimate the (limiting) covariance matrix of the estimator $\hat{\beta}$, by the same methods as in the GLM, thus permitting valid inference. The case of the unrestricted covariance matrix in Eq. (1.42), however, does not allow for consistent estimation of the (limiting) covariance matrix of the estimator $\hat{\beta}$, and thus valid inference in this case is not possible.

REFERENCES

Balestra, Pietro and Marc Nerlove (1966), "Pooling Cross Section Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas", *Econometrica*, pp. 585-612.

Dhrymes, P.J. (1969) "Alternative Asymptotic Tests of Significance and Related Aspects of 2SLS and 3SLS Estimated Parameters", *Review of Economic Studies*, vol. 36, pp. 213-226.

Dhrymes, P. J. (1978), *Introductory Economics*, New York: Springer-Verlag,

Dhrymes, P. J. (1989) *Topics in Advanced Econometrics: Probability Foundations*, New York: Springer-Verlag.

Dhrymes, P. J. (2000), *Mathematics for Econometrics*, third edition, New York: Springer-Verlag.